

Untitled

PAT-NO: JP355159245A
DOCUMENT-IDENTIFIER: JP 55159245 A
TITLE: OUTPUT METHOD OF JAPANESE LANGUAGE INFORMATION
PUBN-DATE: December 11, 1980

INVENTOR-INFORMATION:

NAME
HARA, TATSUJI
KIMURA, HISATADA
TAKANO, RIKUO

INT-CL (IPC): G06F003/16, G06F003/02 , G06K015/00 , G06F015/38

US-CL-CURRENT: 704/3, 704/205 , 704/258 , 715/534

ABSTRACT:

PURPOSE: To enable the audio output, Kana (Japanese syllabary) output and Kanji (Chinese character) output, without increasing the file capacity, by representing Japanese information with Kana and storing the homonym with additional information specified.

CONSTITUTION: The information processing unit is shown in 1, and Japanese language information A provides the Kana Aa representing the reading and the symbol Ab representing the additional information, and the audio output and the Kana output are made with the audio output unit 2 and the Kana output unit 3 by using the Kana Aa. Further, the Kana-Kanji conversion table 4 is retrieved by using the Kana Aa and the symbol Ab, and the Kanji and Kana mixed sentence corresponded is output from the Kanji output unit 5 with Furigana (Kana letters of Japanese phonetic syllabary), if required.

COPYRIGHT: (C)1980,JPO&Japio

⑩ 日本国特許庁 (JP)

⑪ 特許出願公開

⑫ 公開特許公報 (A)

昭55-159245

⑬ Int. Cl.³

識別記号

庁内整理番号

⑭ 公開 昭和55年(1980)12月11日

G 06 F 3/16

7629-5B

3/02

7010-5B

G 06 K 15/00

7629-5B

// G 06 F 15/38

7201-5B

発明の数 1

審査請求 有

(全 3 頁)

⑮ 日本語情報出力方法

横須賀市武1丁目2356番地日本
電信電話公社横須賀電気通信研
究所内

⑯ 特 願 昭54-66161

⑰ 出 願 昭54(1979)5月30日

⑱ 発 明 者 高野陸男

⑲ 発 明 者 原辰次

横須賀市武1丁目2356番地日本
電信電話公社横須賀電気通信研
究所内

横須賀市武1丁目2356番地日本
電信電話公社横須賀電気通信研
究所内

⑳ 出 願 人 日本電信電話公社

㉑ 発 明 者 木村久正

㉒ 代 理 人 弁理士 小林将高

明 細 書

1. 発明の名称

日本語情報出力方法

2. 特許請求の範囲

漢字、ひらがな、カタカナなどで表記される日本語情報を読みを表わす仮名と同音語を一義的に定める付加情報を表わす記号とを用いて情報処理装置へ格納しておき、前記情報処理装置から音声出力装置を用いて音声出力、仮名出力装置を用いて仮名出力、仮名-漢字変換ファイルと漢字出力装置を用いて仮名漢字振り出力の少なくとも1つを出力することを特徴とする日本語情報出力方法。

3. 発明の詳細な説明

この発明は、漢字、ひらがな、カタカナなどで表記される日本語情報を仮名ベースで格納し、音声出力あるいは印字出力するようにした日本語情報出力方法に関するものである。

従来の日本語情報の格納・出力方法には、カタカナと漢字ベースの方法の2種類があった。

カタカナベースの方法は、日本語情報をその読

みをカタカナで表記し格納する方法で、音声出力、仮名出力が可能で、文字の出現頻度が明らかになっており、その頻度が分野にほとんどよらないため情報圧縮が容易にできるという利点はあるが、漢字出力をするためにはカタカナ-漢字変換の技術を利用する必要があり、変換処理が大変である。同音語が存在するため一義的に漢字変換できないなどの問題点が存在するため、漢字出力には適していないという欠点があった。

一方、漢字ベースの方法は、日本語情報をその表記のまま格納する方法で、漢字出力には適しているが、漢字の読みは一義的でないので、音声出力、仮名出力には適していない。また、漢字の出現頻度は明確になっていないばかりでなく分野によつて大きく異なるため情報圧縮が容易にできないという欠点があった。

したがって、音声出力と仮名出力、漢字出力を行うためには、カタカナベースのファイルと漢字ベースのファイルの両方を備えておく必要があり、ファイル容量が大きくなるという欠点があった。

(1)

(2)

この発明はこれらの欠点を除去するためになされたもので、日本語情報を仮名で表記するとともに同音語に対しては所定の付加情報を付加して格納することによりファイル容量を増加させることなく、音声出力、仮名出力、漢字出力を可能としたものである。以下この発明について説明する。なお、この明細書においては、仮名は平仮名と片仮名を意味し、平仮名の場合には「かな」または「ひらがな」と、片仮名の場合には「カナ」または「カタカナ」と表わすことにする。

第1図はこの発明の一実施例を示すブロック図である。1は情報処理装置で日本語情報Aは、その読みを表わす仮名Aa、付加情報表わすある記号Abとで格納されており、仮名Aaを用いて音声出力および仮名出力を音声出力装置2および仮名出力装置3よりそれぞれ行う。さらに、仮名Aaと記号Abを用いて仮名-漢字変換テーブル4を検索し、該当する漢字仮名混り文を漢字出力装置5から、必要ならばふり仮名付きで出力する。日本語情報Aの格納方法と仮名-漢字変換テ-

(3)

なお、付加情報Abを表わす記号としては、同音語を一義的に決定できるものならば何でもよく、同音語内のシーケンシャル番号や、本発明者らが先に提案した特徴識別記号、例えば、漢字の形態の特徴により縦割リ(例: 田)、横割リ(例: 田)の記号田、田等が考えられる。

以上説明したようにこの発明は、読み仮名に付加情報を付加した形で日本語情報を情報処理装置へ格納し出力する方法であるので、次のような利点がある。

- (1) 1つのデータベースで必要に応じて音声出力、仮名出力、漢字仮名混り出力(振仮名付き)が可能であるため、データベースの容量を増加させることなく多様な出力要求に応じることができる。
- (2) 従来のカナベースのシステムに付加情報と変換テーブルを加えるだけの簡単な変更で容易に上記要求に応じることができる。
- (3) 仮名ベースの格納であるので情報圧縮が容易にできる。

(5)

ブル4の構成について以下にその一例を示す。仮名-漢字変換での大きな問題は①変換語の抽出と②同音語の識別である。したがって、上記2つの問題点を解決するため、例えば変換語の前接を付加情報表わす記号Abで囲み、その記号Abの識別によつて同音語の識別を行う。その一例を以下に示す。

「漢字を出力する」

①カンジ②ヲ③シユツリヨク④スル

このようにすると、変換語は完全に抽出されており、その読みを表わす仮名Aaと記号Abとを第2図に示す仮名-漢字変換テーブル4で検索すれば、該当する漢字に変換することができる。

すなわち、第2図で、1は見出し部、2は付加情報部、3は漢字部を表わしており、カンジの例として同音語が5種例示されている。これらの同音語は付加情報Abとして、「00」、「01」……等により区別される。そして「シユツリヨク」のように同音語のない場合には「00」を付加情報Abとして加えてもよい。

(4)

4. 図面の簡単な説明

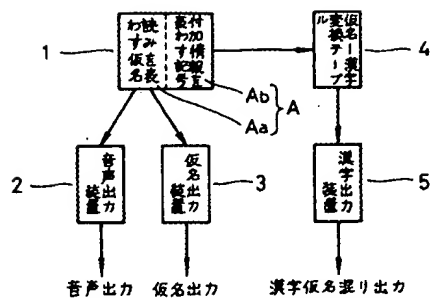
第1図はこの発明の一実施例を示すブロック図、第2図は第1図の仮名-漢字変換テーブルの一例を示す図である。

図中、Aは日本語情報、Aaは読みを表わす仮名、Abは付加情報表わす記号、1は情報処理装置、2は音声出力装置、3は仮名出力装置、4は仮名-漢字変換テーブル、5は漢字出力装置である。

代理人 小林 将 高

(6)

第 1 図



第 2 図

カンジ	0 0	監 事
カンジ	0 1	漢 字
カンジ	0 2	幹 事
カンジ	0 4	充 治
カンジ	0 5	感 じ
シュツリョク	0 0	出 力
シュツルイ	0 0	出 産

見出し部 付加情報部 漢字部